EasyChair Preprint

№ 4795

Analysis of various health parameters for early and efficient prediction of sepsis

Abhishek Parashar, Yukti Mohan and Neeru Rathee

December 25, 2020

Analysis of various health parameters for early and efficient prediction of sepsis

ABHISHEK PARASHAR[1], YUKTI MOHAN[2], AND NEERU RATHEE[3,*]

[1]ECE Department, Maharaja Surajmal Institute of Technology, C-4, janakpuri, New Delhi, India
[2]ECE Department, Maharaja Surajmal Institute of Technology, C-4, janakpuri, New Delhi, India
[3]ECE Department, Maharaja Surajmal Institute of Technology, C-4, janakpuri, New Delhi, India
[*]Corresponding author: neeru1rathee@gmail.com

Compiled March 16, 2020

Sepsis is a major health issue causing mortality, morbidity and health care financial crisis to people around the globe. To resolve this issue, many researchers and clinical practitioners have worked hard to predict the onset of sepsis using various parameters of patients. The proposed work is an attempt of authors to analyse the various parameters (8 vital parameters, 26 laboratory or clinical parameters, and 6 demographic parameters regarding hospital stay) given in Physionet Challenge dataset so as to devise the best features for early and efficient prediction of sepsis. Authors have also addressed another important issue of missing values of some parameters of some patients by applying Gaussian Mixture Model to estimate the missing value in pre-processing steps. The pre-processed data is then fed to Extreme Gradient Boosting algorithm(XGBoost), which is a state of the art performer algorithm for prediction purposes in data analysis field. The experimental results show that by real time monitoring of data from cloud, sepsis can be predicted 6 hours prior to the onset of sepsis with an accuracy score of 0.994 and AUC score of 0.867. It is also observed that demographic parameters play a vital role in sepsis prediction. The Since the parameters used for early prediction can be easily acquired with the help of sensors, the proposed approach proves its potential for development of mobile and website applications for patient monitoring, real-time prediction of sepsis and generation of appropriate alert system.

keywords: Sepsis, Gaussian Mixture Model, Health Parameters, XGBoost, Real Time Analysis

http://dx.doi.org/10.1364/ao.XX.XXXXXX

1. INTRODUCTION

Sepsis is a medical state that occurs when body's response to infections causes tissue damage, organ failure, or death [1]. It is a major health issue around the globe causing mortality and morbidity along with financial burden [2] [3]. In U.S.A alone approximately 750,000 patients suffer from severe sepsis and mortality rate corresponds to approximately 33 percent of this amount [4]. The cost incurred in sepsis treatment is exorbitant, an average cost of US$20.3 billion annually, or US$55.6 million per day in US hospitals is estimated [5, 6]. Additionally, there is a 13 percent increase in average severe sepsis indices [7]. Early and unerring treatment of sepsis can reduce effects by substantial amounts whereas, delayed treatment results in an increase of mortality rate approximately by 4-8 % with each passing hour [8] [9] [10]. Machine learning methods have helped in the detection and prediction of sepsis to a large extent utilizing laboratory parameters and other vital indications but, real-time and precise early detection of sepsis before the critical state is reached is still an issue. Representing sepsis by using clinical parameters is also highly complex and contrastive as sepsis depends on different biased parameters such as genetic variation and immune response state. Additionally, there is a variation in the origin of the infection.

For clinical monitoring of patients derived approaches like Systemic inflammatory response syndrome (SIRS) criteria [11] or sequential organ failure assessment (SOFA) scores [12] are being studied for decades. However, their application is limited due to the heterogeneous nature of sepsis and simplicity trade-off.

Machine learning approaches like Random Forest Classifier, Decision Tree, Gradient Boosting or sequential learning approach like Long Short-Term Memory have helped in solving the issue of early detection of sepsis to a larger extent [13] [14]. A major portion of work on sepsis was done for early detection of sepsis using systematic inflammatory response syndrome (SIRs) criteria using a model named InSight [15]. Some of the researchers focused only on a specific group of patients for example Jin carried out a major work focusing on trauma sepsis patients [16]. Nemati et. al.[17] in their work illustrated a machine learning model for onset prediction of sepsis 4-12 hours before the clinical treatment diagnosis . Similar work was done by Kumar et. al. [18] by proposing a model for prediction of sepsis 4 hours before the onset of sepsis.

The various researchers in the field of sepsis prediction have worked independently by acquiring their own dataset and their own features. Keeping in mind the fact, recently a dataset was released for Physionet Challenge [19]. The researchers have explored the dataset for sepsis prediction by using various algorithms such as signature based model [20], applying Long Short Term Memory (LSTM) on aggregated representations [21], applying autoencoders [22], using random forest classifiers [23], using hybrid features [24] etc. Another important experimentation has been done by [25] by providing a comparison of LSTM and XGBoost for early prediction of sepsis. Almost all the above researchers have worked to increase the prediction time using all the features but the analysis regarding contribution of various features individually for early prediction is missing in the literature. To find the contribution of features towards early prediction of sepsis, authors have analyzed the features of three different categories: vital parameters, laboratory or clinical parameters, and parameters regarding hospital stay. The best AUC score achieved using the proposed approach is 0.867 for 6 hour prior prediction of sepsis. The contribution of the proposed work is 3-fold:

1. The proposed approach analyze the features so as to find which features among vital, laboratory and demographic parameters play vital role for sepsis prediction.

2. There are missing values in the given database. To handle this problem, authors propose to use GMM and catagorical analysis to estimate the missing values.

3. The prediction time achieved using the proposed approach is 6 hours prior to the onset of sepsis with an AUC score of 0.867.

4. The proposed approach is implemented using cloud based servers thus proving its potential for real time monitoring of data for sepsis prediction before its onset. So, it provides an impetus towards web based real time health monitoring for sepsis prediction.

The outline of the proposed work is shown in figure 1. The various parameters of the patients given with the database are first preprocessed to fill the missing values. The preprocessed data is used to train the XGBoost algorithm for early prediction of sepsis. The rest of the paper is organized as follows: section 2 describes the preprocessing step. XGBoost algorithm is described in detail in section 3. The database used to evaluate the proposed approach is pesented in section 4. Experimental settings and results to analyze the various parameters is explained in section 5. Finally, conclusion and future scope is mentioned in section 6.

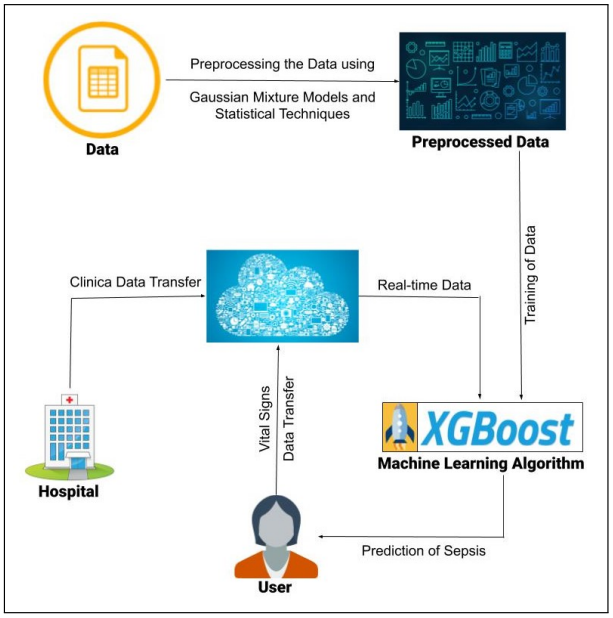

Fig. 1. Workflow of the proposed approach.

2. DATA SET DESCRIPTION

The presented work has been evaluated to analyze the training set of the 2019 PhysioNet/Computing in Cardiology Challenge [19]. The data was accumulated over a decade from over 40,336 patients from 2 distinct hospitals. The longest ICU stay recorded was for 336 hours. The shortest ICU stay accounted for 8 hours. In most of the cases, ICU stay ranged between 20-35 hours. The rate of sepsis increases by 12.5 percent after 60 hours of prolonged ICU stay. The sepsis definition also changes from hospital-acquired to ICU-acquired sepsis. Thus, these values were removed. Sepsis data for patients having label 1 with less than 6 hours of ICU stay were also removed. The dataset contains 40 distinct parameters which are further categorized as 8 vital parameters, 26 laboratory or clinical parameters, and 6 hospital stay parameters as shown in table 1. Thus, we infer that sepsis depends a lot on ICU admission time and other hospital parameters along with, vital parameters like heart rate, breath rate, etc. There were a lot of missing values in the dataset to make the challenge more robust. To overcome this issue of missing values and other erroneous values feature engineering was applied.

3. PRE-PROCESSING

Pre-processing is the primary step for any data analysis problem. There are a number of preprocessing steps listed in the literature such as normalization, excluding the outliers etc. The database used in the proposed work poses a different problem of missing values of some parameters for some patients to check the robustness of the proposed algorithm. To deal with this issue, authors have used Gaussian Mixture model for imputation since the given dataset fits well in a Gaussian distribution. The GMM a soft clustering technique proposed by [26] and used by many researchers for missing value imputation in IoT field [27].

Table 1. Various parameters of PhysioNet Challenge 2019 Dataset

Vital Signs		Laboratory Values			
HR	Heart rate (beats per minute)	BaseExcess	Measure of excess bicarbonate (mmol/L)	Glucose	Serum glucose (mg/dL)
O2Sat	Pulse oximetry (%)	HCO3	Bicarbonate (mmol/L)	Lactate	Lactic acid (mg/dL)
Temp	Temperature (Deg C)	FiO2	Fraction of inspired oxygen (%)	Magnesium	(mmol/dL)
SBP	Systolic BP (mm Hg)	pH	N/A	Phosphate	(mg/dL)
MAP	Mean arterial pressure (mm Hg)	PaCO2	Partial pressure of carbon dioxide from arterial blood (mm Hg)	Potassium	(mmol/L)
DBP	Diastolic BP (mm Hg)	SaO2	Oxygen saturation from arterial blood (%)	$Bilirubin_total$	Total bilirubin (mg/dL)
Resp	Respiration rate (breaths per minute)	AST	Aspartate transaminase (IU/L)	Troponinl	Troponin I (ng/mL)
EtCO2	End tidal carbon dioxide (mm Hg)	BUN	Blood urea nitrogen (mg/dL)	Hct	Hematocrit (%)
Demographics		Alkalinephos	Alkaline	Hgb	Hemoglobin
Age	Years (100 for patients 90 or above)	phosphatase (IU/L)	(mg/dL)	(g/dL)	partial thromboplastin time (seconds)
Gender	Female (0) or Male (1)	Chloride	(mmol/L)	WBC	Leukocyte count (count*$10^3/L$)
Unit1	Administrative identifier for ICU unit (MICU)	Creatinine	(mg/dL)	Fibrinogen	(mg/dL)
Unit2	Administrative identifier for ICU unit (SICU)	$Bilirubin_direct$	Bilirubin direct (mg/dL)	Platelets	(count*$10^3/L$)
HospAdmTime	Hours between hospital admit and ICU admit				
ICULOS	ICU length-of-stay (hours since ICU admit)				

A Gaussian Mixture Model is a combination of g Gaussians where g ranges between $\{1 \cdots G\}$. Here G is the maximum number of clusters in a given data set. Each Gaussian g is comprises of the following parameters:

- Mean μ which defines the center of the Gaussian.
- Co-variance σ, which defines the width of the Gaussian.
- Mixing probability π which defines the height of the Gaussian.

These mixing probabilities must fulfill the probability criteria that is:

$$\sum_{g=1}^{G} \pi_g = 1 \quad \textbf{(1)}$$

The Gaussian Density function is given as:

$$K(x|\mu,\sigma) = \frac{1}{(2\pi)^{T/2}}|\sigma^{1/2} \exp\left(\frac{-1}{2}(x-\mu)^D\sigma^{-1}(x-\mu)\right) \quad \textbf{(2)}$$

The above equation maybe rewritten as:

$$ln(x|\mu,\sigma) = -\frac{T}{2}\ln 2\pi - \frac{1}{2}\ln\sigma - \frac{1}{2}(x-\mu)^D\sigma^{-1}(x-\mu) \quad \textbf{(3)}$$

where x represents a data point, T represents the dimensions of each data point, μ and σ represent mean and co-variance respectively. This is called the *'Maximum Likelihood'* algorithm and the solution of this equation helps to define the optimal values of the Gaussian parameters(μ, σ, π) called the Maximum Likelihood Estimates (MLE).

The percentage of missing values was computed for each column to analysis the missing value data and represented in figure 2. For the final model, all the columns having more than 90% missing values were dropped. To apply GMM to the given data, each column was divided into two halves, thus making 2 clusters. The missing value in these columns was then replaced by the mean μ of the best fitted Gaussian. The distribution plot of various parameters is shown in figure 3 to prove the efficacy of the GMM on the given data. Since ensemble techniques work best on categorical data, the values of the given parameters were then converted into categories.

The correlation plot of various parameters play a vital role in presenting the dependency of parameters with each other. The correlation plot with and without missing values is shown in the figures 4 and 5 to prove the obliteration of some parameters from the given data. Parameters which have been categorized are not included in the correlation plot as shown in figure 5 as the correlation of categorical values holds no relevance.

4. XGBOOST ALGORITHM

XGBoost is a type of ensemble Machine Learning algorithm being popularly used in data analysis community for various

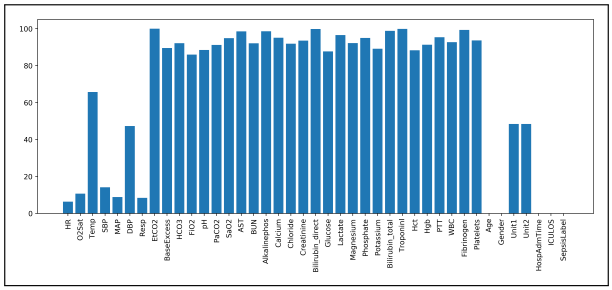

Fig. 2. Percentage of missing values for each parameter.

prediction purposes. XGBoost is based on the sequential structure of decision trees and makes use of the properties of gradient boosting. XGBoost has proven to be useful in many classification and regression problems, hence covering a wide range of applications. As mentioned in the introduction section XGBoost performs even better than LSTM and deep networks. Being inspired from the promising results, XGBoost have been implemented for sepsis prediction in the proposed approach.

XGboost was first proposed by [28] and it has been a favorite of the research community since then. The success of XGBoost is credited to its scalability in all scenarios. Gradient boosting algorithms have also been proposed in previous studies presented by [29] and [30], but XGBoost surpasses all these algorithms and is de facto the best machine learning algorithm for larger datasets, simply because it is a scalable algorithm for tree boosting.

It is a machine learning algorithm, which has an inverted tree like structure, where each internal node represents a feature and each leaf node represents a class. It mimics the functioning of a flowchart in classification and regression problems. Applying bootstrap aggregating or bagging on decision trees results in the formation of Random Forest which is a combination of a large number of uncorrelated independent models. Bagging is a type of machine learning ensemble meta-algorithm which is designed to improve the stability and accuracy of various other machine learning algorithms for classification and regression problems. Boosting refers to the process in which the result of the model is based on the learning from the previous predictions. Gradient boosting is a special boosting technique in which one tree is built at a time and each new tree works to reduce the errors made by the previous trained trees. This led to the formation of another algorithm called Extreme Gradient Boosting or XGBoost.

The loss function of the XGBoost Algorithm is defined as:

$$J^{(m)} = \sum_{i=1}^{n} j(y_i, \hat{y}_i^{(m-1)} + f_m(x_i)) + \theta(f_m) \quad \textbf{(4)}$$

where y_i is the Label from the training data set and $\hat{y}_i^{(m-1)} + f_m(x_i)$ can be observed as $f(a + \Delta a)$ where a can be observed as $\hat{y}_i^{(m-1)}$.

XGBoost utilizes the *Taylor Approximation* to transform the original loss function into a Euclidean Domain function, so that traditional optimization techniques can be applied on it.

After applying the second order Taylor approximation, the modified loss function is given as follows:

$$\tilde{L}^{(m)} = \sum_{i=1}^{n} [s_i f_m(x_i) + \frac{1}{2} r_i f_m^2(x_i)] + \theta(f_m) \quad \textbf{(5)}$$

where, s_i and r_i are the first and second gradient statistics of the loss function.

Since in XGBoost, every newly build tree helps to improve the errors of the previous trained trees, a *next learner* needs to be built.

For this purpose, we need to find the optimal next learner, which can be obtained using the quality scoring function which is defined as:

$$\tilde{L}^{(m)}(q) = \frac{-1}{2} \sum_{k=1}^{M} \frac{(\sum_i \in I_k s_i)^2}{\sum_i \in I_k r_i + \rho} + \gamma M \quad \textbf{(6)}$$

This quality scoring function or 'q' function returns the minimum value of the loss for the given learner.

1. Start with the root.
2. Traverse all the features and their values and evaluate loss for each split as: **gain** = **loss**(farther instances) – (**loss**(left branch) + **loss**(right branch))
3. The gain for the best split must be positive otherwise we must stop growing the branch.

This algorithm is called *'Exact Greedy Algorithm'*. It's complexity is given by **O(x*y)** where x is the number of training samples and y is the dimensions of the features.

5. RESULTS AND DISCUSSIONS

Data streaming pipeline was developed and the proposed model was uploaded to the cloud. We evaluated the model for sepsis detection 6 hours prior to the onset of sepsis. An accuracy score of 0.994 was achieved. Since accuracy score is not a sole parameter for judging the performance of a model, ROC curve and AUC score were calculated. The AUC score after pre-processing and applying Gaussian mixture model is 0.867 whereas the AUC score before pre-processing and applying Gaussian mixture model is 0.743. This shows that pre-processing and replacing missing values with Gaussian values have a profound effect on the performance of the model. The ROC curve for model trained on data without pre-processing and pre-processed data is shown in the figures 7 and 8 respectively. A sharp elbow in the ROC curve shows that the model is working exceedingly well and is following ideal conditions for ROC curve.

By studying various parameters and their effect on the performance of the model, we find that hospital stay parameters or demographic data determine the onset of sepsis to a great extent. An AUC score of 0.942 was achieved with demographic parameters which shows that onset of sepsis is largely dependent on demographic parameters and these parameters can be used for earliest detection of sepsis. For another model trained on vital parameters gives an AUC score of 0.674 and an AUC score of 0.634 was achieved for clinical or laboratory parameters which show that onset is least dependent on these parameters. We also analysed joint effect of different parameters. When demographic parameters are combined with vital parameters, an AUC score of 0.707 is achieved. Similarly, an AUC score of 0.678 is achieved by combining demographic parameters with laboratory parameters. An AUC score of 0.639 is achieved by combining vital parameters and laboratory parameters. The ROC curves for various analysis has been shown in the figure 9.

On comparing different machine learning algorithms like XG boost, ADA boost and Random forest classifier, maximum AUC score was achieved with XG boost algorithm, thus, justifying its use for the presented approach. A comparison table of different

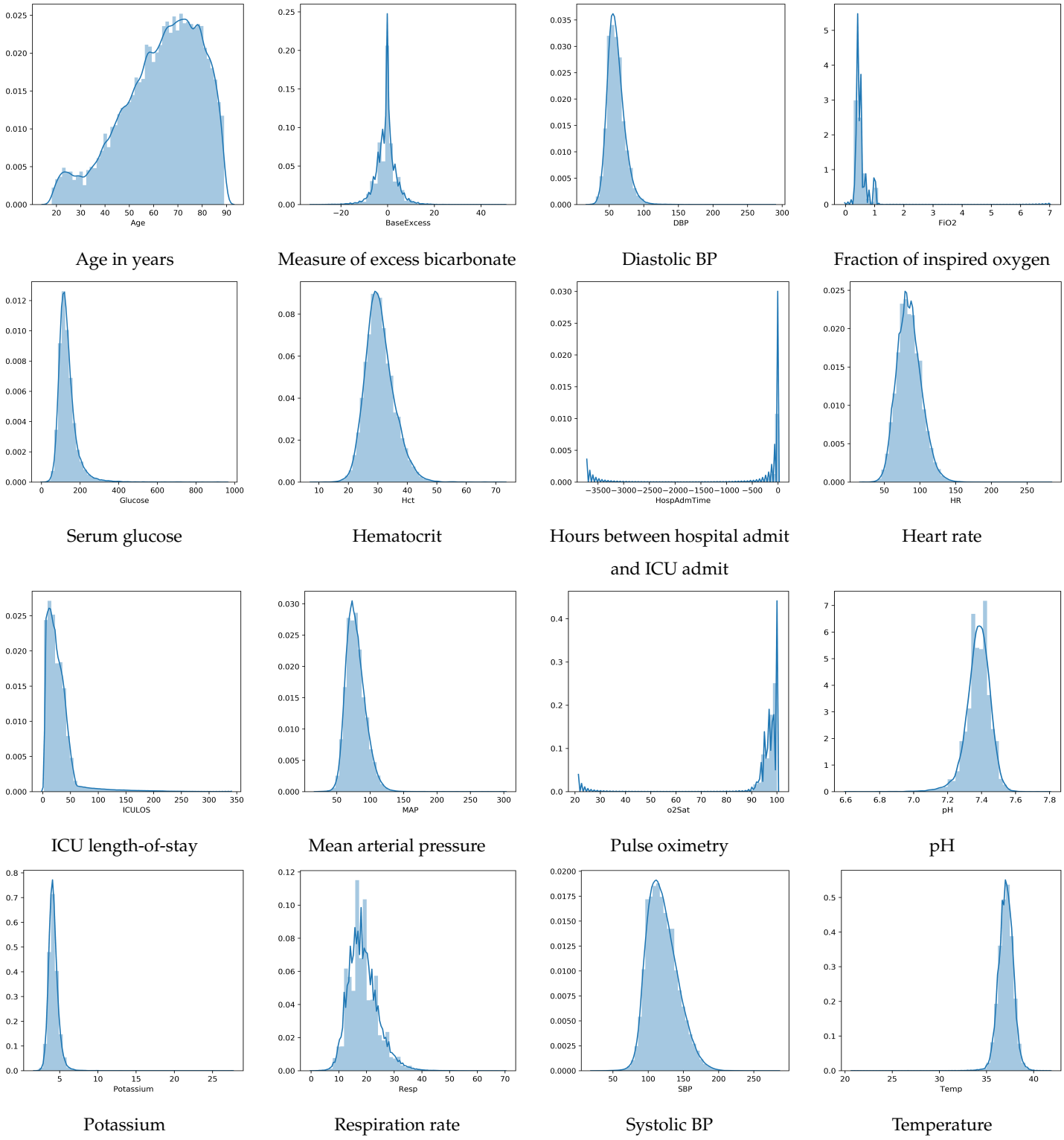

Fig. 3. Density plots of selected parameters.

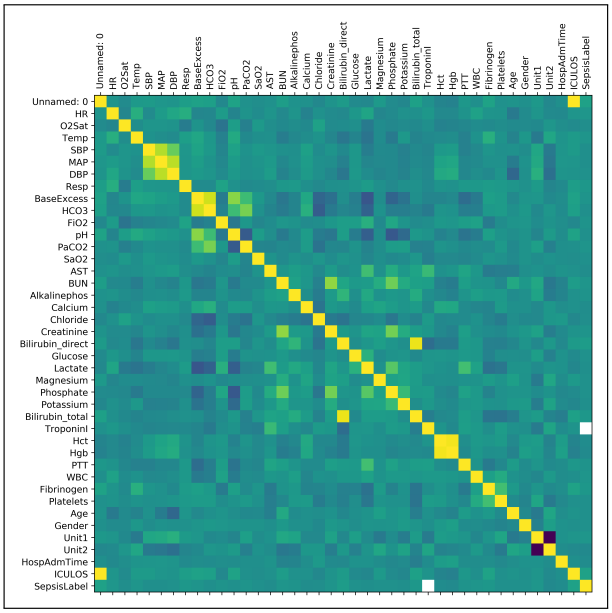

Fig. 4. Correlation plot of every given parameter.

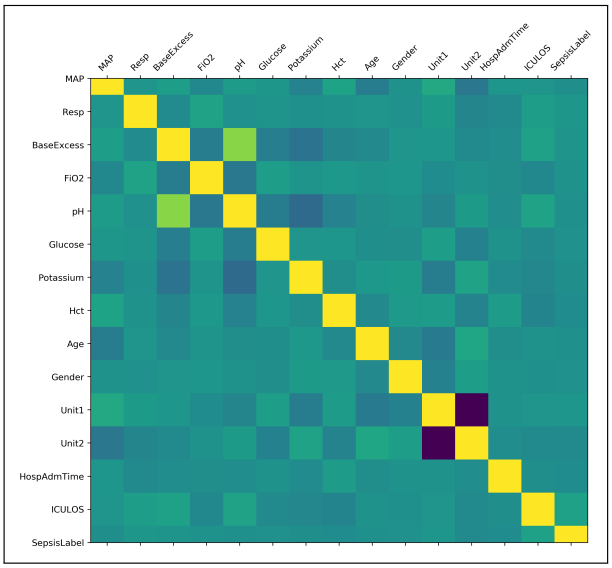

Fig. 5. Correlation plot of selected parameters.

Fig. 6. Evolution of XGBoost.

machine learning algorithm is also given in the table 2. Comparing the performance of our model with some of the pre-existing works as shown in table 3 show the rightness of our model.

Therefore, the proposed model is efficient, accurate, fast and robust in detecting sepsis 6 hours before the onset. Since the model is deployed on the cloud and works with real-time data, streamlined through special apache beam pipelines, it can be deployed in a clinical setting to detect sepsis automatically. As

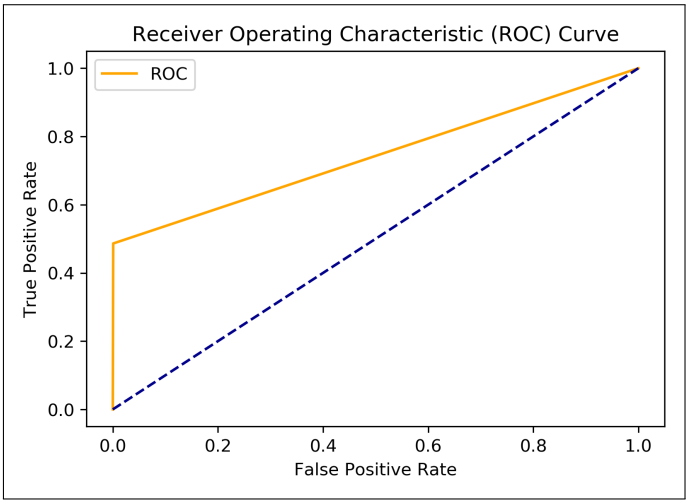

Fig. 7. ROC curve for model trained on raw data.

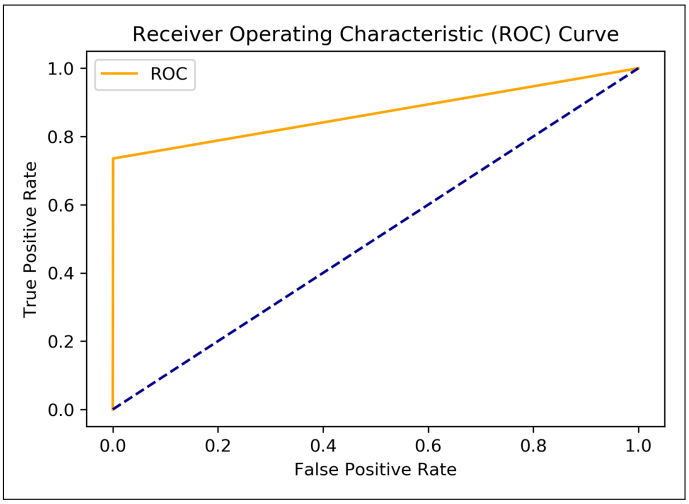

Fig. 8. ROC curve for model trained on pre-processed data.

the model is trained on multiple parameters and diverse patients from different hospitals, it would work fine with distinct patients. In the future, a larger and more diverse dataset can be used for training of the model and overall accuracy can be improved further.

6. CONCLUSION

Real-time and precise prediction of sepsis prior to the onset was a major problem. We proposed a solution to this issue with our work. Our model is accurate, fast and precise to a greater extent and predicts sepsis way before the onset of the deadly disease. Our model is deployed on google cloud platform for real-time analysis and apache beam, an open-source software for creating data streaming pipelines, is used for transferring real-time data. Since pre-processing is of great importance when dealing with clinical data, we used GMM to get best fit for filling the missing values. As a future scope of this project, we propose the development of mobile and web-based applications and systems for real-time monitoring of the patients. These applications can also be useful in collecting data and sending alert messages to the concerned authorities thus, completely automating the process of sepsis detection. The model accuracy

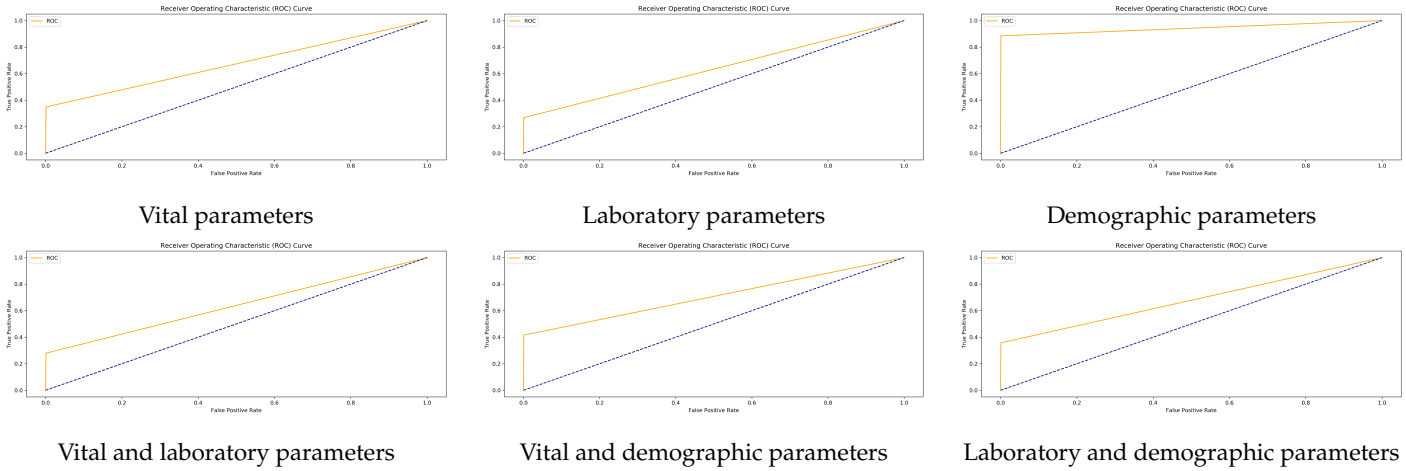

Fig. 9. Receiver operating characteristics of various features alone and in combination with each other.

AUC Score		
1	Vital Parameters	0.67
2	Laboratory Parameters	0.63
3	Demographic Parameters	0.94
4	Vital and Laboratory	0.64
5	Laboratory and Demographics	0.68
6	Demogaphics and Vital	0.71
7	All (without GMM) using XG Boost	0.74
8	**All (with GMM) using XG Boost**	**0.867**
9	All (with GMM) using ADA Boost	0.67
10	All (with GMM) using Random Forest	0.50

Table 2. AUC Score using various features alone and in combination with each other

and precision can be further increased by using more data sets and extensive training. Also, several deep learning models for sequential learning can be applied to make the model more robust.

Study	Algorithm used	AUC Score
Morrill et al. [20]	Signature-based model	0.868
Macias et al. [22]	LSTM RNN	0.79
Lyra et al. [23]	Random forest	0.81
Vollmer et al. [31]	XGBoost	0.823
Fu et al. [32]	Ensemble model	0.792
Firoozabadi et al. [33]	Ensemble model	0.764
Luan et al. [34]	Auto encoders	0.81
Liu et al. [21]	LSTM	0.841
Proposed approach	XGBoost (Without GMM)	0.74
Proposed approach	**XGBoost (with GMM)**	**0.867**

Table 3. Comparison of AUC scores of various studies in 2019.

	Positive prediction	**Negative prediction**
Positive class	65575	37
Negative class	368	1022

Table 4. Confusion Matrix for proposed approach.

F1 Score	Recall	Precision	Accuracy
0.9967	0.9994	0.994	0.994

Table 5. Performance Metrics for proposed approach.

REFERENCES

1. M. Singer, C. S. Deutschman, C. W. Seymour, M. Shankar-Hari, D. Annane, M. Bauer, R. Bellomo, G. R. Bernard, J.-D. Chiche, C. M. Coopersmith *et al.*, "The third international consensus definitions for sepsis and septic shock (sepsis-3)," Jama **315**, 801–810 (2016).
2. D. C. Angus, W. T. Linde-Zwirble, J. Lidicker, G. Clermont, J. Carcillo, and M. R. Pinsky, "Epidemiology of severe sepsis in the united states: analysis of incidence, outcome, and associated costs of care," Critical care medicine **29**, 1303–1310 (2001).
3. M. Moss, "Epidemiology of sepsis: race, sex, and chronic alcohol abuse," Clin. infectious diseases **41**, S490–S497 (2005).
4. E. K. Stevenson, A. R. Rubenstein, G. T. Radin, R. S. Wiener, and A. J. Walkey, "Two decades of mortality trends among patients with severe sepsis: a comparative meta-analysis," Critical care medicine **42**, 625 (2014).
5. A. Pfuntner, L. M. Wier, and C. Steiner, "Costs for hospital stays in the united states, 2010: statistical brief# 146," Healthc. cost utilization project (HCUP) statistical briefs (2006).
6. J. O'Brien, "The cost of sepsis," CDC Safe Healthc. Blog (2015).
7. D. F. Gaieski, J. M. Edwards, M. J. Kallan, and B. G. Carr, "Benchmarking the incidence and mortality of severe sepsis in the united states," Critical care medicine **41**, 1167–1174 (2013).

8. H. B. Nguyen, E. P. Rivers, B. P. Knoblich, G. Jacobsen, A. Muzzin, J. A. Ressler, and M. C. Tomlanovich, "Early lactate clearance is associated with improved outcome in severe sepsis and septic shock," Critical care medicine **32**, 1637–1642 (2004).
9. A. Kumar, D. Roberts, K. E. Wood, B. Light, J. E. Parrillo, S. Sharma, R. Suppes, D. Feinstein, S. Zanotti, L. Taiberg *et al.*, "Duration of hypotension before initiation of effective antimicrobial therapy is the critical determinant of survival in human septic shock," Critical care medicine **34**, 1589–1596 (2006).
10. C. W. Seymour, F. Gesten, H. C. Prescott, M. E. Friedrich, T. J. Iwashyna, G. S. Phillips, S. Lemeshow, T. Osborn, K. M. Terry, and M. M. Levy, "Time to treatment and mortality during mandated emergency care for sepsis," New Engl. J. Medicine **376**, 2235–2244 (2017).
11. F. Jaimes, J. Garcés, J. Cuervo, F. Ramírez, J. Ramírez, A. Vargas, C. Quintero, J. Ochoa, F. Tandioy, L. Zapata *et al.*, "The systemic inflammatory response syndrome (sirs) to identify infected patients in the emergency room," Intensive care medicine **29**, 1368–1371 (2003).
12. J.-L. Vincent, R. Moreno, J. Takala, S. Willatts, A. De Mendonça, H. Bruining, C. Reinhart, P. Suter, and L. G. Thijs, "The sofa (sepsis-related organ failure assessment) score to describe organ dysfunction/failure," (1996).
13. H. J. Kam and H. Y. Kim, "Learning representations for the early detection of sepsis with deep neural networks," Comput. biology medicine **89**, 248–255 (2017).
14. T. Van Steenkiste, J. Ruyssinck, L. De Baets, J. Decruyenaere, F. De Turck, F. Ongenae, and T. Dhaene, "Accurate prediction of blood culture outcome in the intensive care unit using long short-term memory neural networks," Artif. intelligence medicine **97**, 38–43 (2019).
15. J. S. Calvert, D. A. Price, U. K. Chettipally, C. W. Barton, M. D. Feldman, J. L. Hoffman, M. Jay, and R. Das, "A computational approach to early sepsis detection," Comput. biology medicine **74**, 69–73 (2016).
16. H. Jin, Z. Liu, Y. Xiao, X. Fan, J. Yan, and H. Liang, "Prediction of sepsis in trauma patients," Burn. & trauma **2**, 106–113 (2014).
17. S. Nemati, A. Holder, F. Razmi, M. D. Stanley, G. D. Clifford, and T. G. Buchman, "An interpretable machine learning model for accurate prediction of sepsis in the icu." Critical care medicine **46**, 547–553 (2018).
18. S. P. Shashikumar, M. D. Stanley, I. Sadiq, Q. Li, A. Holder, G. D. Clifford, and S. Nemati, "Early sepsis detection in critical care patients using multiscale blood pressure and heart rate dynamics," J. electrocardiology **50**, 739–743 (2017).
19. M. A. Reyna, C. S. Josef, R. Jeter, S. P. Shashikumar, M. B. Westover, S. Nemati, G. D. Clifford, and A. Sharma, "Early prediction of sepsis from clinical data: the physionet/computing in cardiology challenge 2019," Critical Care Medicine (2019).
20. J. Morrill, A. Kormilitzin, A. Nevado-Holgado, S. Swaminathan, S. Howison, and T. Lyons, "The signature-based model for early detection of sepsis from electronic health records in the intensive care unit," in *2019 Computing in Cardiology Conference (CinC). IEEE,* (2019).
21. L. Liu, H. Wu, Z. Wang, Z. Liu, and M. Zhang, "Early prediction of sepsis from clinical datavia heterogeneous event aggregation," arXiv preprint arXiv:1910.06792 (2019).
22. E. Macias, G. Boquet, J. Serrano, J. Vicario, J. Ibeas, and A. Morell, "Novel imputing method for the early prediction of sepsis in icu using deep learning techniques," .
23. S. Lyra, S. Leonhardt, and C. H. Antink, "Early prediction of sepsis using random forest classification for imbalanced clinical data," in *2019 Computing in Cardiology (CinC),* (IEEE, 2019), pp. 1–4.
24. J. Yao, M. L. Ong, K. K. Mun, S. Liu, and M. Motani, "Hybrid feature learning using autoencoders for early prediction of sepsis," in *2019 Computing in Cardiology (CinC),* (IEEE, 2019), pp. Page–1.
25. Y. Wang, B. Xiao, X. Bi, W. Li, J. Zhang, and X. Ma, "Prediction of sepsis from clinical data using lstm and xgboost," .
26. D. A. Reynolds, "Gaussian mixture models." Encycl. biometrics **741** (2009).
27. X. Yan, W. Xiong, L. Hu, F. Wang, and K. Zhao, "Missing value imputation based on gaussian mixture model for the internet of things," Math. Probl. Eng. **2015** (2015).
28. T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining,* (2016), pp. 785–794.
29. J. Ye, J.-H. Chow, J. Chen, and Z. Zheng, "Stochastic gradient boosted distributed decision trees," in *Proceedings of the 18th ACM conference on Information and knowledge management,* (2009), pp. 2061–2064.
30. S. Tyree, K. Q. Weinberger, K. Agrawal, and J. Paykin, "Parallel boosted regression trees for web search ranking," in *Proceedings of the 20th international conference on World wide web,* (2011), pp. 387–396.
31. M. Vollmer, C. F. Luz, P. Sodmann, B. Sinha, and S.-O. Kuhn, "Time-specific metalearners for the early prediction of sepsis," in *2019 Computing in Cardiology (CinC),* (IEEE, 2019), pp. 1–4.
32. M. Fu, J. Yuan, M. Lu, P. Hong, and M. Zeng, "An ensemble machine learning model for the early detection of sepsis from clinical data," in *2019 Computing in Cardiology (CinC),* (IEEE, 2019), pp. Page–1.
33. R. Firoozabadi and S. Babaeizadeh, "An ensemble of bagged decision trees for early prediction of sepsis," in *2019 Computing in Cardiology (CinC),* (IEEE, 2019), pp. Page–1.
34. M. N. Luan Tran and C. Shahabi, "Representation learning for early sepsis prediction," Computing. **46**, 1 (2019).